Assignment 6 Solution

CS 432

Timothy Filippone

```
Problem 1.  Find 3 users who are closest to you in terms of age,
gender, and occupation.  For each of those 3 users:

- what are their top 3 favorite films?
- bottom 3 least favorite films?

Based on the movie values in those 6 tables (3 users X (favorite +
least)), choose a user that you feel is most like you.  Feel
free to note any outliers (e.g., "I mostly identify with user 123,
except I did not like ``Ghost'' at all").
```

Solution:

The function def ClosestUsers(amount,identity) reads u.user and compares the given identity list to users in the file. Once a set of user are found, the highest and lowest ratings amongst those users are taken.

I mostly identify with user 66. I have seen movies like Goodfellas such as Reservoir Dogs(1992). Except, I would not enjoy watching DragonHeart. I also relate to the user's bottom three favorites.

Top 3 Favorite:

User		Rating	Time	Title
66	181	5	883601425	GoodFellas (1990)
66	471	5	883601296	Dragonheart (1996)
66	742	5	883601388	Crow: City of Angels, The (1996)
37	403	5	880915942	Pinocchio (1940)
37	56	5	880915810	Priest (1994)
37	161	5	880915902	On Golden Pond (1981)
33	313	5	891963290	3 Ninjas: High Noon At Mega Mountain (1998)
33	328	4	891964187	Desperate Measures (1998)
33	343	4	891964344	Apostle, The (1997)

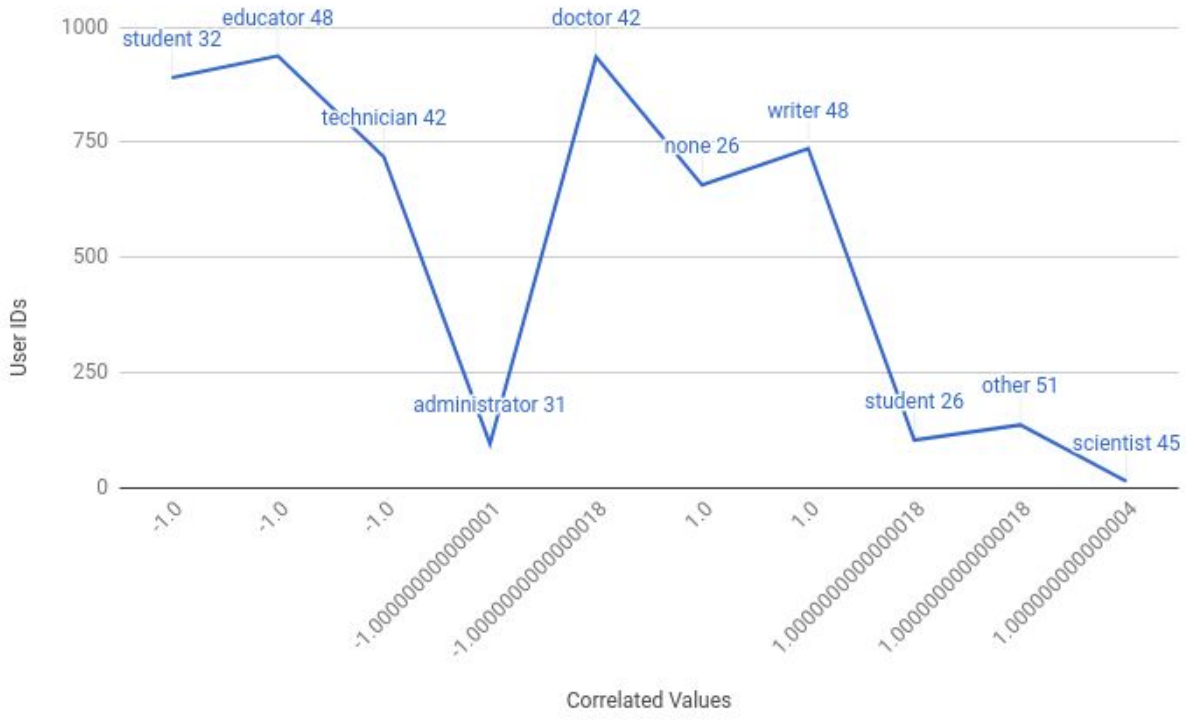
Bottom 3 Favorite:

User		Rating	Time	Title
66	877	1	883601089	That Darn Cat! (1997)
66	21	1	883601939	Braveheart (1995)
66	286	1	883601089	Marvin's Room (1996)
37	82	1	880915942	Much Ado About Nothing (1993)
37	540	2	880916070	Mortal Kombat (1995)
37	825	2	880915565	Phantom, The (1996)
33	872	3	891964230	Picture Perfect (1997)
33	307	3	891964148	FairyTale: A True Story (1997)
33	895	3	891964187	Sweet Hereafter, The (1997)

Problem 2. Which 5 users are most correlated to the substitute you? Which 5 users are least correlated (i.e., negative correlation)?

Solution:

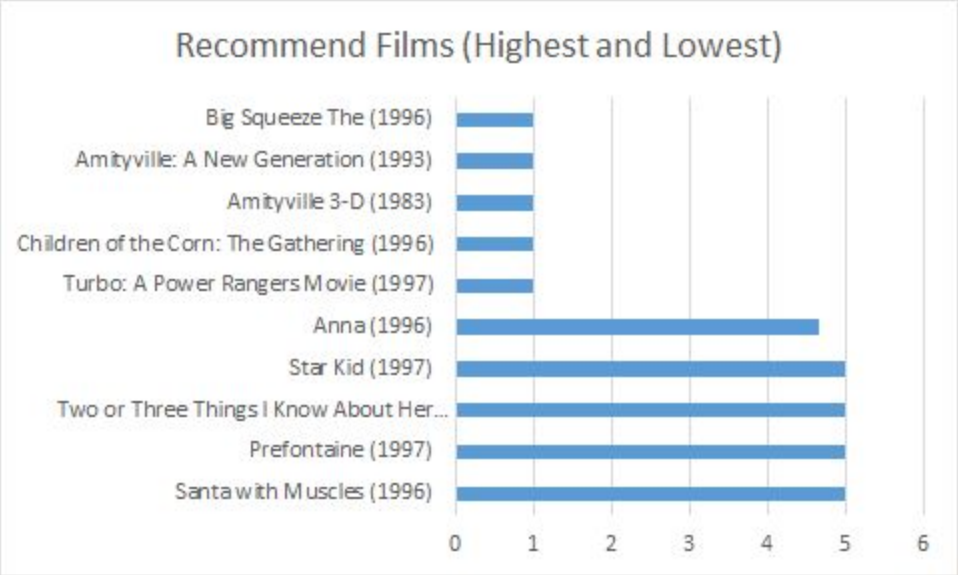
To gather the correlated users, def compareUsers(prefs,p1) is used. Users are stored in a list with sorted by their correlation rank. This value is set using sim_pearson(prefs, p1, p2) function.



Problem 3. Compute ratings for all the films that the substitute you have not seen. Provide a list of the top 5 recommendations for films that the substitute you should see. Provide a list of the bottom 5 recommendations (i.e., films the substitute you is almost certain to hate).

Solution:

A recommendation list is taken from getRecommendations(prefs,p1) and acquired to def notSeen(prefs,p1). Movies already seen by the user are removed from the recommended list. Ratings from u.data are added to each film as well. Recommend films are then sorted from highest to lowest.



Problem 4. Choose your (the real you, not the substitute you) favorite and
least favorite film from the data. For each film, generate a list
of the top 5 most correlated and bottom 5 least correlated films.

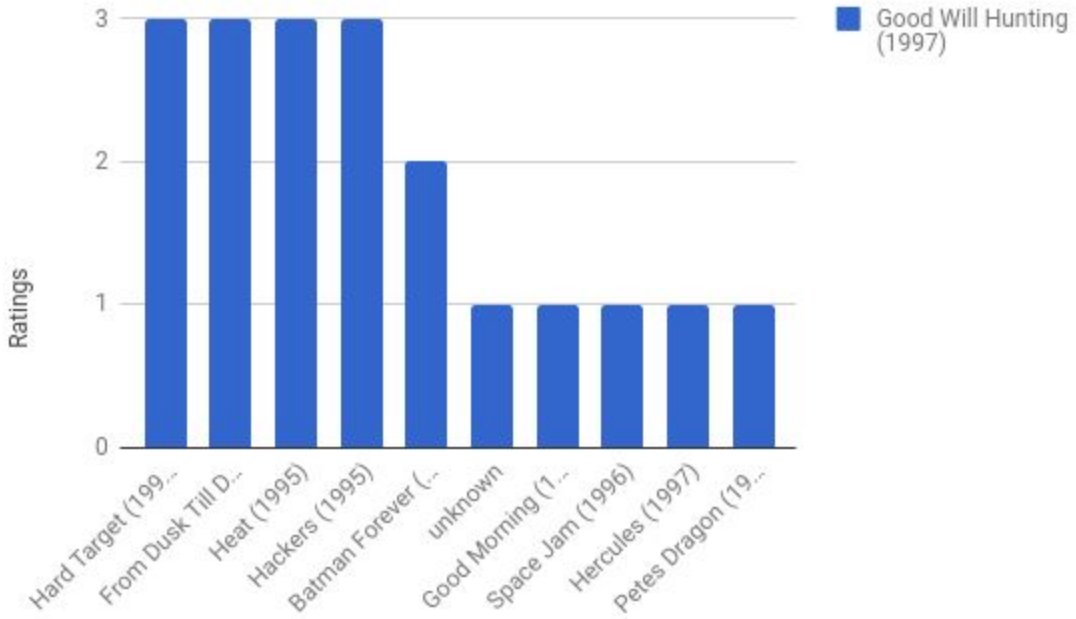
Based on your knowledge of the resulting films, do you agree with
the results? In other words, do you personally like / dislike
the resulting films?

Solution:

def maxminRecommendedFilms(filmID) sorts a list of related films for a user given film ID number. Based on that film's genre binary string, other films from u.item are compared. Each item is given a count number based on how matches to the film's genre binary string, depending on highest for ones or lowest for zeros.

I would say the results for my favorite film are close to what I prefer. The lowest results like Space Jam(1996 and Hercules(1997) are relative to my disliking. As for the film I most dislike, the results do not seem accurate to my taste. The Big Lebowski (1998) is a film I do like. However, I did notice Good Morning (1971) appear in both graphs.

Good Will Hunting (1997)



The Wings of the Dove (1997)